Digital Humanities, Social Science and Cultural Preservation ISSN: 3069-2237

PRIME OPEN ACCESS

Volume 2, Issue 3

Research Article

Date of Submission: 20 May, 2026
Date of Acceptance: 15 June, 2026
Date of Publication: 16 July, 2026

A Conceptual Framework for Artificial Intelligence Combining Buddhism and the Free Energy Principle

Yasunari Miyagi[1,2]*, Yoko Tateishi[2], Moe Yorozu[2] and Takehi Yamagata[2]

[1]Medical Data Labo, Okayama city, Japan

[2]Department of Obstetrics and Gynecology, Sun Clinic, Okayama city, Japan

***Corresponding Author:** Yasunari Miyagi, MD, PhD., Medical Data Labo, Naka ward, Okayama, Okayama prefecture 703-8267, Japan

Citation: Miyagi.Y., Tateishi.Y., Yorozu.M., Yamagata.T. (2026). A Conceptual Framework for Artificial Intelligence Combining Buddhism and the Free Energy Principle. *Digit Hum Soc Sci Cult Preserv, 2*(3), 01-05.

Abstract

Remarkable advancements in artificial intelligence (AI) have been made in recent years. There is a theory that brain activity in humans is based on the free energy principle, which is a comprehensive theory that attempts to explain the structure and function of cognition and the brain's behavior. For any self-organized system to be in equilibrium within its environment, the agent performs active inference to minimize the system's informational free energy and reduce the prediction error from the current presumption. On the other hand, the theory based on the vijñapti-mātratā of Buddhist philosophy explains consciousness through a multilayered structure comprising manifestation and information storage and seems to result in a certain degree of compatibility with AI systems that perform calculations and store information. Therefore, an AI that integrates the vijñapti-mātratā philosophy with the free energy principle could achieve performance, enabling us to perceive not only knowledge but also actions that include human-like emotional expressions. We hereby report that an AI system theory that integrates the theory of vijñapti-mātratā with the principle of free energy.

Keywords: Artificial Intelligence, Buddhism, Consciousness, Free Energy Principle, Model, Vijñapti-Mātratā

Introduction

Remarkable advancements in artificial intelligence (AI) have been made in recent years. AI continuously accumulates data while performing statistical pattern matching computationally, thereby improving its accuracy. Technologies for recognizing and classifying image information, along with semantic segmentation, have been developed, leading to the creation of AI-powered autonomous driving systems for vehicles [1,2]. Generative AI powered by transformers trained on textual information can be said to have already been put to practical use in tasks such as summarizing text and proposing solutions to problems [3,4]. These systems are indeed based on mimicking human neural networks, but it remains unclear whether their computational processes resemble electrical signal processing within the human brain. In particular, generative AI is known to produce hallucinations, and humans still often determine whether AI outputs are truthful. Furthermore, even if we recognize that an AI's output expresses emotion, we have no way to determine whether what is actually generated internally by the AI can be called emotion.

There is a theory that brain activity in humans is based on the free energy principle. The free energy principle with active inference is a comprehensive theory that attempts to explain the structure and function of cognition and the brain's behavior [5-9]. The theory is based on a variational Bayesian estimate to provide a comprehensive explanation of perception, action, emotion, sentiment, and decision-making. For any self-organized system to be in equilibrium within its environment, the agent performs active inference to minimize the system's informational free energy and reduce the prediction error from the current presumption to the generated model. An agent that uses a free energy formulation of perception adjusts its internal states and samples the environment to minimize its free energy. To put it simply, the brain

has a property that causes it to operate in a way that minimizes the difference between the probability distribution of perceptions generated within the brain and the probability distribution of perceptions obtained from the sensory organs. There are reports on the free energy principle concerning adults and human fetuses [10-14].

Furthermore, reports indicate that the free energy principle can be applied even to a single neuron or a fertilized egg [15,16]. Therefore, if this free energy principle is applied to AI, it can be inferred that the output of AI might be similar to that of biological organisms. Research on deep neural networks based on the free energy principle has been reported [17]. Pietarinen et al. (2025) proposed that the free-energy principle as a unifying arbiter that set an explanatory pathway toward integrated cognition and meaning [18]. However, while the free energy principle requires the generation of a probability distribution of perception from perceptual information, it cannot be said with certainty that it determines exactly how perception is formed.

On the other hand, the study of cognition has a long history of research by philosophers since ancient times. Scientific proof has not been established for Kant's idea that "The understanding can intuit nothing; the senses can think nothing. In no other way than from their union can knowledge arise" (1781) Husserl's phenomenology (1936) or Barclay's idealism that "To be is to be perceived" (1713) [19-21]. Among theories of cognition, the theory based on Buddhist philosophy's vijñapti-mātratā that is a soteriological framework in which consciousness is transformed (ālaya is purified and the manas relinquishes its misapprehension of self), proposed in India during the 4th--5th centuries, explains consciousness through a multilayered structure comprising manifestation and information storage [22]. This seems to result in a certain degree of compatibility with AI systems, though AI systems would not possess a soteriological function. According to the vijñapti-mātratā, consciousness arising in the brain from sensory input is generated from the underlying unconscious layer, manas-vijñāna, which is conditioned by each individual's experiences and learning. Beneath this lies alaya-vijñāna, where all unconscious information concerning the organism's life is stored and available. Therefore, an AI that integrates the vijñapti-mātratā philosophy with the free energy principle of neuroscience could achieve performance, enabling us to perceive not only knowledge but also actions that include human-like emotional expressions. We hereby report an AI system theory that integrates the theory of vijñapti-mātratā with the principle of free energy.

Materials and Methods

Vijñapti-Mātratā and the Free Energy Principle

In the vijñapti-mātratā conception, beneath consciousness lies an unconscious layer shaped by each individual's experiences and learning. This layer is defined as manas-vijñāna. Manas-vijñāna is a collective entity composed of numerous unconscious pieces of information concerning the living organism within an individual and a misapprehension of the stream of consciousness itself as a substantive "I." Each element within manas-vijñāna repeatedly undergoes generation, change, and cessation through causality and is not fixed. Furthermore, manas-vijñāna is a subset of alaya-vijñāna, which is said to contain all the unconscious information regarding the organism within an individual. Each element within alaya-vijñāna repeatedly undergoes generation, change, and cessation through causality and is not fixed.

Recognizing the sense of the self begins when information enters the brain through the perceptions that are usually vision, hearing, taste, smell, and touch. Touch is thought to include bathyesthesia. Information from these senses forms a probability distribution of inferred information within the brain, known as recognition density in terms of the free energy principle. Moreover, probably almost simultaneously, according to the vijñapti-mātratā, information from the perceptions is matched within the manas-vijñāna, which is inherent to each individual life form and possesses personal characteristics, to information approximating the newly formed recognition density derived from the perceptions. This matched information is then selected as a model. In addition, according to the original free energy principle, a model is generated from information received through the perceptions, with the resulting probability distribution of this information defined as the generative density. In vijñapti-mātratā, this model is selected from among those generated causally within the manas-vijñāna, whereas in the free energy principle, it is generated each time. Although there is this difference, both share the commonality of requiring the existence of a generated model.

Furthermore, on the basis of the Kullback–Leibler divergence and Bayes' theorem changes occur in accordance with the principle of free energy, moving in the direction that minimizes the variational free energy [23,24]. According to the vijñapti-mātratā interpretation, consciousness is thought to exist when there is a change in the variational free energy, or consciousness is thought to generate biological phenomena that decrease the variational free energy. At this point, you recognize it as an entity possessing knowledge, emotion, thought, and will.

We created a mathematical notation for a system that integrates the above concepts, unifying consciousness on the basis of the theory of vijñapti-mātratā with the principle of free energy. We demonstrated an example.

Results

Formula Combining the Vijñapti-Mātratā and Free Energy Principle

The free energy, F, based on the free energy principle incorporating the theory of vijñapti-mātratā is as follows:

$$F(\tilde{o},\mu) = D_{KL}[Q(\tilde{s},\tilde{u}|\mu)||P(\tilde{s},\tilde{u}|\tilde{o})] - \ln P(\tilde{o}|mdl_{\Omega})$$

$$\mu_t = arg\,min_{\mu}\, F(\{o_0,\ldots,o_{t+1}\},\mu)$$

$$a_t = \arg min_{a} \sum_{\Omega} P(o_{t+1}|o_t,a)\, F(\{o_0,\ldots,o_{t+1}\},\mu_t)$$

$$M = \{mdl_a, mdl_b, mdl_c, \ldots\}$$

$$A = \{M_\alpha, M_\beta, M_\gamma, \ldots\}$$

$$\tilde{o} = (o_1, o_2, \ldots, o_t)$$

$$\tilde{s} = (s_1, s_2, \ldots\ s_t)$$

$$\tilde{u} = (u_1, u_2, \ldots\ u_t)$$

where a is actions, A is alaya-vijñāna, D_{KL} is Kullback–Leibler divergence, mdl is a model, o_t is observations, M is manas-vijñāna, P is generative density, Q is recognition density, s_t is hidden states, u is prediction of the result of causing an action, Ω is a set of observations, μ is sufficient statistics, and μ_t is perception [25].

An example of consciousness and behavior using the theory of vijñapti-mātratā and the free energy principle is shown in Figure. 1.

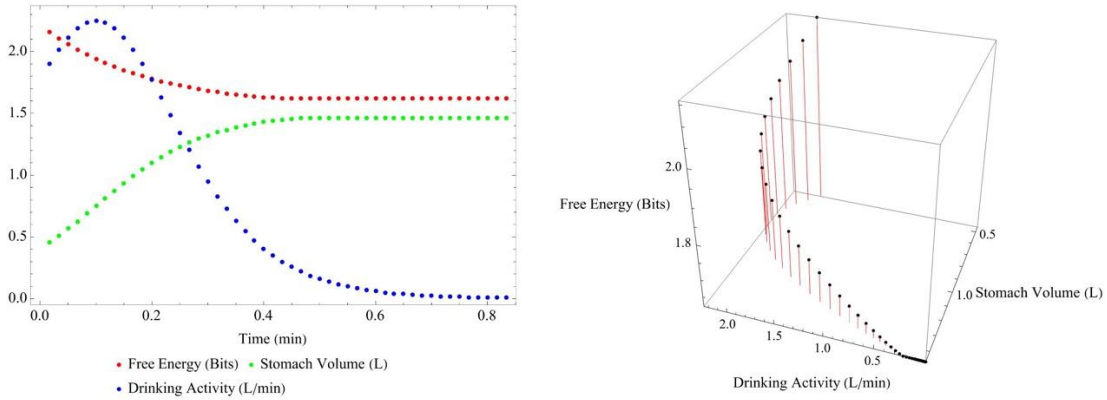

Figure. 1. A case where sensory information indicating elevated plasma osmotic pressure in the brain was received under certain conditions. First, when the plasma osmotic pressure is high, a probability distribution called generative density is formed. Then, on the basis of the learned information model within manas-vijñāna that high plasma osmotic pressure indicates a tendency toward dehydration and necessitates drinking water, a probability distribution called recognition density arises in the brain via empirical Bayesian theory. According to the principle of free energy, if the Kullback–Leibler divergence indicates a difference between the two probability distributions, one becomes aware of a tendency toward dehydration and the need to drink water. As organisms drink water in response to increasing awareness, the water content in the stomach increases, lowering the plasma osmotic pressure and reducing the free energy value, which then stabilizes, prompting the cessation of drinking behavior. In the left panel, chronological variations in the free energy value (bits), stomach volume (L), and drinking activity (L/min) are shown in red, green and blue, respectively. As drinking behavior increases the stomach volume, the free energy decreases and stabilizes. In the right panel, the relationships among the free energy value (bits), stomach volume (L), and drinking activity (L/min) are shown as dots in a three- dimensional graph. The red line represents the free energy value at each point. The interval between dots corresponds to 1s.

Discussion

A mathematical system integrating Buddhist philosophy's vijñapti-mātratā with neuroscience's free energy principle has been presented. Vijñapti-mātratā is not scientifically proven in humans, and the free energy principle is merely one theory within the field of neuroscience. Therefore, this system only demonstrates potential. However, this integrated system might prove useful for developing AI that we can perceive as human-like. Specifically, many models capable of

responding to input information for AI and organizing them into a database should be created. Using the generative density produced by these models corresponding to the input information and the recognition density formed from the input information, the AI should take actions that minimize the variational free energy. This is based on the Kullback–Leibler divergence, Bayesian theory, and the free energy principle. We can infer that by doing so, people would perceive this AI as exhibiting reactions extremely close to those of humans.

Kant stated, "Time and space are not determinations of things in themselves, but only forms of our sensible intuition," asserting that humans perceive phenomena through space and time (1781) [19]. However, AI should be capable of fully understanding phenomena occurring in high-dimensional spaces beyond four-dimensional space. Furthermore, by utilizing various sensors capable of detecting light and electromagnetic waves at wavelengths beyond human visible light, temperatures that are too extreme for human skin to endure, sound waves outside the human audible range, and even stimuli from smell and taste, and feeding these data as inputs to AI, it is possible that in the future, AI systems capable of determining optimal actions beyond human judgment could be developed.

Furthermore, it is an AI capable of being recognized as possessing extremely advanced intelligence, and it has the potential to continue growing toward the future as an asset for all humanity. In other words, the information of models continuously stored and accumulated within the massively generated alaya-vijñāna can persist as a vast database beyond the biological lifespan of humans and the service life of computer storage devices. Wei reported (2026) that agentic reasoning for LLMs that is the paradigm that enables models to autonomously plan, adapt, and verify via multi-step workflows leverages sequential loops of planning, action selection, tool invocation, and context updates to systematically solve complex problems and empirical studies show that combining behavior priming with reinforcement learning significantly improves accuracy and exploratory capability [26,27]. The necessity for appropriate exploration aligns with our perspective.

On the other hand, it is crucial that the models of alaya-vijñāna are good. If the information acquired and accumulated is not appropriately curated, it may lead to adverse consequences for society. While Plato's Theory of Forms is famous for defining what constitutes the good, there has been a vast body of philosophical knowledge on this subject since ancient times [28]. Future AI could impact all living beings and objects beyond humans, so the training of models should be conducted with extreme caution. We believe philosophy and ethics will become even more essential in the future than they are today.

Limitations

First, a scientifically verifiable method to prove the vijñapti-mātratā is still unknown to humanity. Furthermore, it is possible that a better theory regarding the generation of human consciousness may be discovered. We cannot definitively state that vijñapti-mātratā is the absolutely optimal theory. Second, since the free energy principle is merely one theory within the field of neuroscience, this system merely demonstrates its potential. Third, as reported by the International Energy Agency in 2025 creating AI for such systems requires enormous amounts of data and electricity [29]. The development of specialized sensors for acquiring diverse types of perceptual information will also be necessary. Therefore, this system cannot immediately achieve AI. Fourth, while the output from this AI may appear to us as if it possesses consciousness, emotions, the ability to think, and free will, whether it can actually be considered a living being depends on interpretation and requires philosophical judgment. When considering the entire universe, we believe that a biological definition of life has not yet been established. Sixth, if information that transcends human cognitive perception is fed into artificial intelligence, we may become unable to determine the morality or veracity of the output from a scientific, philosophical, or ethical standpoint. Therefore, even for artificial intelligence incorporating vijñapti-mātratā, its development should proceed with caution.

Conclusion

A mathematical system that integrates the consciousness theory of the Buddhist doctrine of vijñapti-mātratā with the free energy principle of neuroscience has been demonstrated. This system may serve as a reference for developing AI that mimics humans.

Statements and declarations

Acknowledgments: None.
Authors' Contributions: All authors contributed to the study conception and design. Material preparation, data collection and analysis were performed by Yasunari Miyagi. The first draft of the manuscript was written by Yasunari Miyagi and all authors commented on previous versions of the manuscript. All authors read and approved the final manuscript.
Funding: None.
Data availability: Not applicable.
Declarations: Conflicts of interest None.
Large Language Models (LLMs): Not used.

References

1. Levinson, J., Askeland, J., Becker, J., Dolson, J., Held, D., Kammel, S., Kolter, J. Z., Langer, D., Pink, O., Pratt, V. R., Sokolsky, M., Stanek, G., Stavens, D., Teichman, A., Werling, M., & Thrun, S. (2011). Towards fully autonomous

driving: Systems and algorithms. 2011 IEEE Intelligent Vehicles Symposium (IV), 163–168.
2. National Highway Traffic Safety Administration. (2020). Automated driving systems. U.S. Department of Transportation.
3. Singh, S. K., Kumar, S., & Mehra, P. S. (2023). Chat GPT & Google Bard AI: A review. In 2023 International Conference on IoT, Communication and Automation Technology (ICICAT) (pp. 1–6). IEEE.
4. Rane, N. L., Choudhary, S. P., & Rane, J. (2024). Gemini versus ChatGPT: Applications, performance, architecture, capabilities, and implementation. Journal of Applied Artificial Intelligence, 5(1), 69–93.
5. Friston, K., Kilner, J., & Harrison, L. (2006). A free energy principle for the brain. Journal of Physiology-Paris, 100(1–3), 70–87.
6. Friston, K. (2009). The free-energy principle: A rough guide to the brain? Trends in Cognitive Sciences, 13(7), 293–301.
7. Friston, K. J., Daunizeau, J., & Kiebel, S. J. (2009). Reinforcement learning or active inference? PLoS ONE, 4(7), e6421.
8. Friston, K. J., Daunizeau, J., Kilner, J., & Kiebel, S. J. (2010). Action and behavior: A free-energy formulation. Biological Cybernetics, 102(3), 227–260.
9. Friston, K. (2010). The free-energy principle: A unified brain theory? Nature Reviews Neuroscience, 11(2), 127–138.
10. Miyagi, Y., Hata, T., & Miyake, T. (2023). Fetal brain activity and the free energy principle. Journal of Perinatal Medicine, 51(7), 925–931.
11. Miyagi, Y., Hata, T., & Miyake, T. (2025). Characteristics of fetal facial expression changes using artificial intelligence: A pilot study. Scientific Reports, 15, 45639.
12. Miyagi, Y., Hata, T., & Miyake, T. (2025). Fetal brain function and artificial intelligence. In A. Kurjak & F. A. Chervenak (Eds.), Donald School textbook of ultrasound in obstetrics and gynecology (5th ed., pp. 732–740). Jaypee Brothers Medical Publishers.
13. Miyagi, Y., Hata, T., Bouno, S., Koyanagi, A., & Miyake, T. (2021). Recognition of facial expression of fetuses by artificial intelligence (AI). Journal of Perinatal Medicine, 49(5), 596–603. https://doi/org/10.1515/jpm-2020-0537
14. Miyagi, Y., Hata, T., Bouno, S., Koyanagi, A., & Miyake, T. (2023). Artificial intelligence to understand fluctuation of fetal brain activity by recognizing facial expressions. International Journal of Gynecology & Obstetrics, 161(3), 877–885.
15. Isomura, T., Shimazaki, H., & Friston, K. J. (2022). Canonical neural networks perform active inference. Communications Biology, 5(1), 55.
16. Miyagi, Y., Mio, Y., Yumoto, K., Hirata, R., Habara, T., & Hayashi, N. (2024). Kinetic energy and the free energy principle in the birth of human life. Reproductive Medicine, 5(2), 65–80.
17. Mazzaglia, P., Verbelen, T., Çatal, O., & Dhoedt, B. (2022). The free energy principle for perception and action: A deep learning perspective. arXiv.
18. Pietarinen, A. V., Beni, M. D., & Shumilina, V. (2025). Developmental bioelectricity as an explanatory framework for cognition and meaning. Biosemiotics, 18, 31–55.
19. Kant, I. (1781/1787). Critique of pure reason. Johann Friedrich Hartknoch.
20. Husserl, E. (1936). Die Krisis der europäischen Wissenschaften und die transzendentale Phänomenologie.
21. Berkeley, G. (1986). Three dialogues between Hylas and Philonous. Open Court Publishing Company.
22. Buswell, R. E., Jr., & Lopez, D. S., Jr. (2014). The Princeton dictionary of Buddhism. Princeton University Press.
23. Joyce, J. M. (2011). Kullback–Leibler divergence. In M. Lovric (Ed.), International encyclopedia of statistical science (pp. 720–722). Springer.
24. Kendall, M., Stuart, A., & Ord, J. K. (1994). The Advanced Theory of Statistics: Distribution Theory. 6th.
25. Inui, T. (2018). The free-energy principle: A unified theory of brain functions. Nihon Shinkei Kairo Zassi, 25, 123-134.
26. Wei, T., Li, T. W., Liu, Z., et al. (2026). Agentic reasoning for large language models. arXiv.
27. Emergent Mind. (2026). Agentic reasoning in large language models.
28. Gerson, L. P. (2023). The idea of the good. In Plato's moral realism (pp. 29–74). Cambridge University Press.
29. IEA. (2025). Energy and AI.